EXHIBIT 12

How Do You Know a Human Wrote This?

Machines are gaining the ability to write, and they are getting terrifyingly good at it.

July 29, 2020

By Farhad Manjoo
Opinion Columnist

To hear more audio stories from publishers like The New York Times, download Audm for iPhone or Android.

I've never really worried that computers might be gunning for my job. To tell the truth, often, I pray for it. How much better would my life be — how much better would *my editor's* life be, to say nothing of the poor readers — if I could ask an all-knowing machine to suggest the best way to start this column? It would surely beat my usual writing process, which involves clawing at my brain with a rusty pickax in the dim hope that a few flakes of wisdom and insight might, like dandruff, settle on the page.

See what I mean? A computer might have helped there. (Like *dandruff*? That's what you're going with, Farhad?) But we writers can be a cocky bunch. Writing is something of an inexplicable trick, and it feels, like telling a joke or making a soufflé, like an inviolably human endeavor.

I've never really worried that a computer might take my job because it's never seemed remotely possible. Not infrequently, my phone thinks I meant to write the word "ducking." A computer writing a newspaper column? That'll be the day.

Opinion | How Do You Know a Human Wrote This? - The New York Times

Well, writer friends, the day is nigh. This month, OpenAI, an artificial-intelligence research lab based in San Francisco, began allowing limited access to a piece of software that is at once amazing, spooky, humbling and more than a little terrifying.

OpenAI's new software, called GPT-3, is by far the most powerful "language model" ever created. A language model is an artificial intelligence system that has been trained on an enormous corpus of text; with enough text and enough processing, the machine begins to learn probabilistic connections between words. More plainly: GPT-3 can read and write. And not badly, either.

> **Sign up for the Opinion Today newsletter** Get expert analysis of the news and a guide to the big ideas shaping the world every weekday morning. Get it sent to your inbox.

Software like GPT-3 could be enormously useful. Machines that can understand and respond to humans in our own language could create more helpful digital assistants, more realistic video game characters, or virtual teachers personalized to every student's learning style. Instead of writing code, one day you might create software just by telling machines what to do.

OpenAI has given just a few hundred software developers access to GPT-3, and many have been filling Twitter over the last few weeks with demonstrations of its surprising capabilities, which range from the mundane to the sublime to the possibly quite dangerous.

To appreciate the potential danger, it helps to understand how GPT-3 works. Language models often need to be trained for specific uses — a customer-service bot used by a retailer might need to be fine-tuned with data about products, while a bot used by an airline would need to learn about flights. But GPT-3 doesn't need much extra training. Give GPT-3 a natural-language prompt — "I hereby resign from Dunder-Mifflin" or "Dear John, I'm leaving you" — and the software will fill in the rest with text that is eerily close to what a human would produce.

Opinion | How Do You Know a Human Wrote This? - The New York Times

These aren't canned responses. GPT-3 is capable of generating entirely original, coherent and sometimes even factual prose. And not just prose — it can write poetry, dialogue, memes, computer code and who knows what else.

GPT-3's flexibility is a key advance. Matt Shumer, the chief executive of a company called OthersideAI, is using GPT-3 to build a service that responds to email on your behalf — you write the gist of what you'd like to say, and the computer creates a full, nuanced, polite email out of your bullet points.

Another company, Latitude, is using GPT-3 to build realistic, interactive characters in text-adventure games. It works surprisingly well — the software is not only coherent but also can be quite inventive, absurd and even funny.

Stew Fortier, a writer, created a zany satire using the software as a kind of muse.

Fortier fed GPT-3 a strange prompt: "Below is a transcript from an interview where Barack Obama explained why he was banned from Golden Corral for life." The system then filled in the rest of the interview, running with the concept that Obama had been banned from an all-you-can-eat buffet.

> **Obama**: Yes. It's true. I am no longer allowed in Golden Corral.
>
> **Interviewer**: Is this because of your extensive shrimp-n-crab legs policy?
>
> **Obama**: Absolutely.
>
> **Interviewer**: What is your extensive shrimp-n-crab legs policy?
>
> **Obama**: Oh, well, in brief, they were offering an all-you-can-eat shrimp-n-crab leg buffet, and I did not hesitate. After I ate so much shrimp and crab that my stomach hurt, I would quietly sneak in and throw more shrimp and crab onto my plate. I did this over and over again until I had cleaned out the buffet and was full of shrimp-n-crab.

Yet software like GPT-3 raises the prospect of frightening misuse. If computers can produce large amounts of humanlike text, how will we ever be able to tell humans and machines apart? In a research paper detailing GPT-3's power, its creators cite

a litany of dangers, including "misinformation, spam, phishing, abuse of legal and governmental processes, fraudulent academic essay writing and social engineering pretexting."

There are other problems. Because it was trained on text found online, it's likely that GPT-3 mirrors many biases found in society. How can we make sure the text it produces is not racist or sexist? GPT-3 also isn't good at telling fact from fiction. "I gave it my own original three sentences about whales, and it added original text — and the way I could tell it was original was that it was pretty much dead wrong," Janelle Shane, who runs a blog called AI Weirdness, told me.

To its credit, OpenAI has put in place many precautions. For now, the company is letting only a small number of people use the system, and it is vetting each application produced with it. The company also prohibits GPT-3 from impersonating humans — that is, all text produced by the software must disclose that it was written by a bot. OpenAI has also invited outside researchers to study the system's biases, in the hope of mitigating them.

These precautions may be enough for now. But GPT-3 is so good at aping human writing that it sometimes gave me chills. Not too long from now, your humble correspondent might be put out to pasture by a machine — and you might even miss me when I'm gone.

Office Hours With Farhad Manjoo

Farhad wants to chat with readers on the phone. If you're interested in talking to a New York Times columnist about anything that's on your mind, please fill out this form. Farhad will select a few readers to call.

The Times is committed to publishing a diversity of letters to the editor. We'd like to hear what you think about this or any of our articles. Here are some tips. And here's our email: letters@nytimes.com.

Follow The New York Times Opinion section on Facebook, Twitter (@NYTopinion) and Instagram.

Farhad Manjoo was a Times Opinion columnist from 2018 to 2023. Before that, they wrote the State of the Art column. They are the author of "True Enough: Learning to Live in a Post-Fact Society."
   @fmanjoo · Facebook

A version of this article appears in print on , Section A, Page 22 of the New York edition with the headline: How Do You Know a Human Wrote This?